Chilean Journal of Statistics
Volume 16, Number 1, June 2025, Pages 60–78
DOI: 10.32372/ChJS.16-01-04

STATISTICAL INFERENCE AND BOOTSTRAPPING
RESEARCH ARTICLE

Fast double bootstrap tests for non-nested beta regressions

FABIO P. LIMA[1] and FRANCISCO CRIBARI-NETO[1,*]

[1]Departamento de Estatística, Universidade Federal de Pernambuco, Recife, Brazil

(Received: 15 May 2025 · Accepted in final form: 30 June 2025)

Abstract

Bootstrap-based inference is investigated for non-nested hypothesis testing in beta regression models with small samples. Specifically, the J and MJ tests are considered, which assess whether one model outperforms its non-nested alternatives by augmenting it with predictors from competing specifications. Their standard bootstrap and fast double bootstrap versions are examined. Monte Carlo simulations reveal that conventional asymptotic tests suffer from size distortions in finite samples. The fast double bootstrap versions of the J and MJ tests yield more accurate inference while requiring minimal additional computational cost. The theoretical results are supported by an empirical application that illustrates the practical value of the proposed methods for model selection in beta regression.

Keywords: Beta regression · bootstrapping · J test · MJ test · non-nested models

Mathematics Subject Classification: Primary 62F40 · Secondary 62J12

1. INTRODUCTION

Two regression models are considered non-nested when neither can be derived from the other by imposing linear restrictions on their parameters. In such cases, likelihood ratio (LR) tests —suitable for nested model comparisons— are not applicable, and alternative procedures must be employed. A prominent solution in the context of linear models is the J test introduced in [1], which constructs an encompassing model that contains all candidate specifications as special cases. The J test evaluates whether including regressors from an alternative model improves the fit of the model under consideration. Under the null hypothesis, the test statistic is asymptotically chi-square (χ^2) distributed.

Despite its widespread use, the J test has important limitations. Its asymptotic justification may yield considerable size distortions in small samples, a concern emphasized by several authors [2, 3]. Furthermore, when multiple competing models are considered, sequential application of the J test increases the risk of inflated type I errors due to repeated testing. To address this limitation, the MJ test, a minimum-statistic-based approach that tests the joint null hypothesis that no model outperforms the others, was proposed in [4]. This strategy enables simultaneous comparison of several non-nested models and reduces the cumulative error associated with multiple pairwise tests.

---

[*] Corresponding author. Email: francisco.cribari@ufpe.br (F. Cribari-Neto)

ISSN: 0718-7912 (print)/ISSN 0718-7920 (online)

http://www.soche.cl/chjs

The extension of these tests to beta regressions introduces additional challenges. Beta regression is a flexible class of models suited for continuous responses restricted to the standard unit interval, with widespread applications in economics, medicine, ecology, and other fields. Its specification includes separate submodels for the mean and precision parameters, which allows modeling heteroskedasticity and varying dispersion [5, 6]. In applied research, analysts often face the need to compare non-nested beta regression models differing not only in included covariates but also in their choice of link functions or in the structure of the precision submodel. As such, non-nested comparisons in this context are common and nontrivial. The J and MJ testing frameworks were extended to beta regressions in [7], where it is demonstrated that bootstrap implementations can improve performance in small samples. Related developments were also proposed for broader generalized additive models for location, scale and shape, as discussed in [8] using results from [9].

Importantly, many empirical studies using beta regression are based on relatively small sample sizes —particularly in economics and the social sciences— making reliable inference difficult. In such cases, bootstrap methods are not only useful but often necessary to achieve reasonable finite-sample accuracy. However, conventional bootstrap methods still fall short in some scenarios. A full double bootstrap offers improved performance but at the cost of a computational burden that is often prohibitive.

Several studies have explored the fast double bootstrap (FDB) method and its applications. The FDB method was employed in [10] for bias correction. A higher-order iterative framework was developed in [11] to enhance the method, and the performance of FDB-based LR tests in beta regression models under nested hypotheses was assessed [12]. Additional applications are discussed in [13, 14, 15, 16]. The present study contributes to this literature by introducing novel FDB implementations of the J and MJ tests, designed to enhance inferential accuracy while maintaining computational efficiency.

Specifically, this article contributes to the literature by introducing FDB versions of the J and MJ tests tailored for beta regression models. The FDB method, originally proposed in [17], provides size improvements comparable to the full double bootstrap while requiring far fewer resampling steps. Its effectiveness was demonstrated in [18] for linear regressions and later in [12] in the context of nested beta regression models.

We advance this research by adapting the FDB method to non-nested hypothesis testing in beta regressions. Our proposed tests are particularly useful in applied settings where competing models differ in link functions, covariates, or precision structures. Through Monte Carlo simulations, we show that the FDB versions offer improvements in finite-sample accuracy compared to their asymptotic and standard bootstrap counterparts. These gains, while sometimes marginal relative to the standard bootstrap, come at a low computational cost —requiring only a single second-level resample per bootstrap replication. Such improvements are especially valuable in empirical applications where the bootstrap p-value is close to the significance threshold and small refinements in accuracy can meaningfully influence model selection decisions. The proposed tests therefore fill an important methodological gap, providing practitioners with a more reliable and computationally feasible tool for model comparison in beta regressions, particularly in small-sample contexts where traditional approaches often exhibit substantial size distortions.

We present an application to model the proportion of income spent on food as a function of household income and the number of people in the household. In this context, we identify a model that outperforms commonly used specifications by employing non-nested hypothesis testing combined with the FDB method. This application illustrates the practical value of the proposed bootstrap framework for empirical model selection. By enabling more reliable inference in the comparison of competing, non-nested specifications, the method contributes to more accurate conclusions in applied research —particularly in settings where standard model selection tools may lack power or exhibit size distortions.

The rest of the article is organized as follows. Section 2 reviews the structure of beta regression models. In Section 3, we present the J and MJ tests adapted for beta regressions. In Section 4, the FDB method is introduced and its implementation is outlined in our setting. Section 5 reports the results of Monte Carlo simulations comparing the performance of various test versions. In Section 6, we illustrate the practical advantages of the proposed methods through an empirical application. In Section 7, our conclusions are stated.

2. THE BETA REGRESSION MODEL

The class of beta regression models introduced in [5] is widely employed for modeling continuous response variables that assume values strictly within the open unit interval $(0,1)$—such as rates and proportions. Let Y be a random variable following a beta distribution with mean $\mu \in (0,1)$ and precision parameter $\phi > 0$, denoted by $Y \sim \mathrm{Beta}(\mu,\phi)$, under the parametrization proposed in [5]. Its probability density function is given by

$$f(y;\mu,\phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1}, \quad 0<y<1, 0<\mu<1, \phi>0,$$

where Γ is the gamma function. Under this parametrization, the expectation and variance of Y are $\mathbb{E}(Y)=\mu$ and $\mathbb{V}\mathrm{ar}(Y)=\mu(1-\mu)/(1+\phi)$, respectively. The precision parameter ϕ governs the dispersion of the distribution: higher values of ϕ correspond to lower variance, conditional on μ.

Consider a sequence of n independent random variables $Y_1,\dots,Y_n$, where each Y_t follows a beta distribution with mean μ_t and precision ϕ_t. The varying precision beta regression model takes the form given by

$$Y_t \sim \mathrm{Beta}(\mu_t,\phi_t), \quad g(\mu_t) = \sum_{i=1}^{k} x_{ti}\beta_i = \boldsymbol{x}_t^\top\boldsymbol{\beta} = \eta_t, \quad h(\phi_t) = \sum_{j=1}^{m} z_{tj}\gamma_j = \boldsymbol{z}_t^\top\boldsymbol{\gamma} = \delta_t,$$

where $\boldsymbol{\beta}\in\mathbb{R}^k$ and $\boldsymbol{\gamma}\in\mathbb{R}^m$ are vectors of unknown parameters, $\boldsymbol{x}_t = (x_{t1},\dots,x_{tk})^\top$ and $\boldsymbol{z}_t = (z_{t1},\dots,z_{tm})^\top$ are vectors of fixed covariates $(k+m<n)$, η_t and δ_t are linear predictors, and $g\colon (0,1)\to\mathbb{R}$ and $h\colon \mathbb{R}_+\to\mathbb{R}$ are strictly monotonic and twice differentiable link functions.

Parameter estimation is carried out via the maximum likelihood method. The corresponding log-likelihood function is given by

$$\ell(\boldsymbol{\beta},\boldsymbol{\gamma}) = \sum_{t=1}^{n}[\log(\Gamma(\phi_t)) - \log(\Gamma(\mu_t\phi_t)) \\ - \log(\Gamma((1-\mu_t)\phi_t)) + (\mu_t\phi_t-1)\log(y_t) + ((1-\mu_t)\phi_t-1)\log(1-y_t)].$$

The score functions for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are given by

$$\boldsymbol{U}_{\boldsymbol{\beta}}(\boldsymbol{\beta},\boldsymbol{\gamma}) = \boldsymbol{X}^\top\boldsymbol{V}\boldsymbol{T}(\boldsymbol{y}^*-\boldsymbol{\mu}^*), \quad \boldsymbol{U}_{\boldsymbol{\gamma}}(\boldsymbol{\beta},\boldsymbol{\gamma}) = \boldsymbol{Z}^\top\boldsymbol{H}\boldsymbol{a},$$

where $\boldsymbol{X}$ is an $n\times k$ matrix and $\boldsymbol{Z}$ is an $n\times m$ matrix whose tth rows are $\boldsymbol{x}_t^\top$ and $\boldsymbol{z}_t^\top$, respectively. The matrices $\boldsymbol{T}$, $\boldsymbol{H}$, and $\boldsymbol{V}$ are diagonal with entries given by $\boldsymbol{T} = \mathrm{diag}\{1/g'(\mu_1),\dots,1/g'(\mu_n)\}$, $\boldsymbol{H} = \mathrm{diag}\{1/h'(\phi_1),\dots,1/h'(\phi_n)\}$, $\boldsymbol{V} = \mathrm{diag}\{\phi_1,\dots,\phi_n\}$, where primes denote derivatives. The vectors $\boldsymbol{y}^* = (y_1^*,\dots,y_n^*)^\top$ and $\boldsymbol{\mu}^* = (\mu_1^*,\dots,\mu_n^*)^\top$, with $y_t^* = \log(y_t/(1-y_t))$ and $\mu_t^* = \psi(\mu_t\phi_t) - \psi((1-\mu_t)\phi_t)$, where ψ is the digamma function, defined as $\psi(w) = \mathrm{d}\log(\Gamma(w))/\mathrm{d}w$. The vector $\boldsymbol{a} = (a_1,\dots,a_n)^\top$ has entries given by $a_t = \partial\ell_t(\mu_t,\phi_t)/\partial\phi_t$, where $\ell_t(\mu_t,\phi_t)$ denotes the contribution of observation t to the log-likelihood function.

The maximum likelihood estimator of $(\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$, denoted $(\widehat{\boldsymbol{\beta}}^\top, \widehat{\boldsymbol{\gamma}}^\top)^\top$, is obtained by solving the system of equations given by $\boldsymbol{U}_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \boldsymbol{0}$ and $\boldsymbol{U}_{\boldsymbol{\gamma}}(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \boldsymbol{0}$. As these equations generally do not admit closed-form solutions, the parameter estimates $(\widehat{\boldsymbol{\beta}}^\top, \widehat{\boldsymbol{\gamma}}^\top)^\top$ are obtained numerically by maximizing the log-likelihood function. Efficient algorithms such as the Broyden-Fletcher-Goldfarb-Shanno (BFGS) quasi-Newton method are typically employed for this task; see [19, 20] for further details.

Asymptotic normality of the maximum likelihood estimator is well established under standard regularity conditions, such as independent observations, differentiable and one-to-one link functions, and a nonsingular Fisher information matrix; see, for instance, [21] for a general treatment in the context of generalized linear models. While analogous results have not been formally derived for beta regression models with varying precision, we assume that similar conditions hold. Under this assumption, we have approximately

$$\begin{pmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{\gamma}} \end{pmatrix} \sim \text{Normal}_{k+m}\left(\begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\gamma} \end{pmatrix}, \boldsymbol{K}^{-1} \right),$$

where $\boldsymbol{K}^{-1}$ is the inverse of the Fisher information matrix, which has the block structure

$$\boldsymbol{K}^{-1} = \boldsymbol{K}^{-1}(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \begin{pmatrix} \boldsymbol{K}^{\boldsymbol{\beta}\boldsymbol{\beta}} & \boldsymbol{K}^{\boldsymbol{\beta}\boldsymbol{\gamma}} \\ \boldsymbol{K}^{\boldsymbol{\gamma}\boldsymbol{\beta}} & \boldsymbol{K}^{\boldsymbol{\gamma}\boldsymbol{\gamma}} \end{pmatrix},$$

with components given by

$$\begin{aligned} \boldsymbol{K}^{\boldsymbol{\beta}\boldsymbol{\beta}} &= \left(\boldsymbol{X}^\top \boldsymbol{V}\boldsymbol{W}\boldsymbol{X} - \boldsymbol{X}^\top \boldsymbol{C}\boldsymbol{T}\boldsymbol{H}\boldsymbol{Z}(\boldsymbol{Z}^\top \boldsymbol{D}\boldsymbol{Z})^{-1}\boldsymbol{Z}^\top \boldsymbol{H}\boldsymbol{T}\boldsymbol{C}^\top \boldsymbol{X} \right)^{-1}, \\ \boldsymbol{K}^{\boldsymbol{\beta}\boldsymbol{\gamma}} &= \left(\boldsymbol{K}^{\boldsymbol{\gamma}\boldsymbol{\beta}} \right)^\top = -\boldsymbol{K}^{\boldsymbol{\beta}\boldsymbol{\beta}} \boldsymbol{X}^\top \boldsymbol{C}\boldsymbol{T}\boldsymbol{H}\boldsymbol{Z}(\boldsymbol{Z}^\top \boldsymbol{D}\boldsymbol{Z})^{-1}, \\ \boldsymbol{K}^{\boldsymbol{\gamma}\boldsymbol{\gamma}} &= (\boldsymbol{Z}^\top \boldsymbol{D}\boldsymbol{Z})^{-1}[\boldsymbol{I}_m + (\boldsymbol{Z}^\top \boldsymbol{H}\boldsymbol{T}\boldsymbol{C}^\top \boldsymbol{X})\boldsymbol{K}^{\boldsymbol{\beta}\boldsymbol{\beta}} \boldsymbol{X}^\top \boldsymbol{C}\boldsymbol{T}\boldsymbol{H}\boldsymbol{Z}(\boldsymbol{Z}^\top \boldsymbol{D}\boldsymbol{Z})^{-1}]. \end{aligned}$$

Here, $\boldsymbol{I}_m$ is the $m \times m$ identity matrix, and $\boldsymbol{W}$, $\boldsymbol{C}$, and $\boldsymbol{D}$ are diagonal matrices whose tth diagonal elements, for $t = 1, \ldots, n$, are given, respectively, by $w_t = \phi_t\,(\psi'(\mu_t\phi_t) + \psi'((1-\mu_t)\phi_t))\,1/(g'(\mu_t))^2$, $c_t = \phi_t\,(\psi'(\mu_t\phi_t)\mu_t - \psi'((1-\mu_t)\phi_t)(1-\mu_t))$, $d_t = \psi'(\mu_t\phi_t)\mu_t^2 + \psi'((1-\mu_t)\phi_t)(1-\mu_t)^2 - \psi'(\phi_t)$, where ψ' is the trigamma function. The inverse Fisher information matrix is used to compute standard errors for the maximum likelihood estimators in beta regressions. These standard errors correspond to the square roots of the diagonal elements of $\boldsymbol{K}^{-1}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}})$.

3. NON-NESTED HYPOTHESIS TESTING IN BETA REGRESSION MODELS

Beta regression models in the class of varying precision include separate regression structures for the mean and precision parameters. Consequently, non-nested beta regression models may differ in their choice of regressors or link functions in either submodel. Let us consider the problem of testing N competing non-nested beta regression models.

For each model M_i, with $i = 1, \ldots, N$, we write M_i: $g(\boldsymbol{\mu}_i) = \boldsymbol{\eta}_i = \boldsymbol{X}_i\boldsymbol{\beta}_i$ and $h(\boldsymbol{\phi}_i) = \boldsymbol{\delta}_i = \boldsymbol{Z}_i\boldsymbol{\gamma}_i$, where $\boldsymbol{\mu}_i$ and $\boldsymbol{\phi}_i$ are vectors containing the mean and precision parameters, respectively; $\boldsymbol{\eta}_i$ and $\boldsymbol{\delta}_i$ are the linear predictors; $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ are model-specific design matrices; and $\boldsymbol{\beta}_i$ and $\boldsymbol{\gamma}_i$ are vectors of regression coefficients. The links g and h, previously defined for scalar arguments, are applied here componentwise to $\boldsymbol{\mu}_i$ and $\boldsymbol{\phi}_i$.

The J test, originally proposed in [1] for linear regression models and later extended to beta regression in [7], is suitable for testing non-nested hypotheses. Let $d_{m(i)}$ and $d_{p(i)}$ denote the number of models that differ from model M_i in the mean and precision submodels, respectively, and define the set of candidate models as $\mathcal{M} \in \{1, \ldots, N\}$.

To test a given candidate model M_l, with $l \in \mathcal{M}$, using the J test, one first estimates the parameters of all competing models M_j, for each $j \in \mathcal{M} \setminus \{l\}$, via the maximum likelihood method.

Next, the $d_{m(l)}$ estimated predictors $\widehat{\boldsymbol{\eta}}_j$ from models that differ in the mean specification are incorporated into the mean submodel of M_l. Similarly, the $d_{p(l)}$ estimated predictors $\widehat{\boldsymbol{\delta}}_j$ from models that differ in the precision specification are included in the precision submodel of M_l. The J test statistic is the LR statistic used to jointly test the exclusion of all $\widehat{\boldsymbol{\eta}}_j$ and $\widehat{\boldsymbol{\delta}}_j$ from the augmented model. Note that this entails including $d_{m(l)}$ and $d_{p(l)}$ additional covariates in the mean and precision submodels, respectively. Model M_l is rejected at the significance level α if $J_l > \chi^2_{1-\alpha, d_{m(l)}+d_{p(l)}}$. Intuitively, model M_l is not rejected if the inclusion of the estimated predictors from the competing models does not improve its fit. Otherwise, it is rejected in favor of at least one competing alternative. However, this has a drawback: when testing a large number of non-nested models, a sequential testing procedure is required in which each null hypothesis —asserting that the mth model is correctly specified— is tested individually. Moreover, the J test may yield no rejections, suggesting that all models are compatible with the data, or it may reject all competing models at once, indicating misspecification across the board. To address this, the MJ test was proposed in [4] to assess whether the true model is included among the set of candidates, thereby avoiding sequential testing and reducing potential size distortions.

The MJ test is conducted by computing J_i for all $i \in \mathcal{M}$ and defining the test statistic as $\mathrm{MJ} = \min\{J_i \colon i \in \mathcal{M}\}$. The null hypothesis that the true model is included in the candidate set is rejected if $\mathrm{MJ} > \chi^2_{1-\alpha,\tau_i}$, where τ_i denotes the degrees of freedom of the J test corresponding to the selected model, that is, the model for which $J_i = \mathrm{MJ}$.

Unlike in linear regression models, beta regression models involve two distinct submodels —one for the mean and another for the precision. Non-nested models may differ in the specification of the mean submodel, the precision submodel, or both. Consequently, the individual J tests involved in the MJ procedure may be based on different numbers of restrictions. This contrasts with the linear regression setting, in which such differences typically do not arise. If the null hypothesis is not rejected, the model corresponding to the minimum J_i value is selected as the most plausible model.

In small samples, the null distributions of the J and MJ statistics may be approximated by their asymptotic counterpart, the χ^2 distribution. Simulation studies have shown that both tests often exhibit distorted rejection rates in such settings. This pattern has been consistently reported in [2, 3, 4]. To mitigate this issue, these authors recommend using a bootstrap version of the J test in linear regression models. In the context of beta regression, this approach was extended in [7] by proposing bootstrap versions of both the J and MJ tests. Simulation results demonstrate that these bootstrap-based procedures yield more accurate finite-sample performance compared to their asymptotic counterparts.

To illustrate the use of bootstrap resampling in the context of the J and MJ tests, let $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top$ denote a vector of independent beta-distributed random variables and $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$ their observed values. Suppose there are two competing non-nested varying precision beta regression models that differ in the regressors included in their mean submodels. Let $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ denote the corresponding design matrices for the mean submodels, and $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$ the associated linear predictors.

Let $\boldsymbol{Z}$ and $\boldsymbol{\delta}$ denote the design matrix and the linear predictor, respectively, for the precision submodel, which is assumed to be the same in both models. The models can be written as M_i: $g(\boldsymbol{\mu}_i) = \boldsymbol{\eta}_i = \boldsymbol{X}_i\boldsymbol{\beta}_i$, $h(\boldsymbol{\phi}_i) = \boldsymbol{\delta_i} = \boldsymbol{Z}\boldsymbol{\gamma}_i$, for $i = 1, 2$.

The bootstrap J test for assessing model M_1 is implemented using Algorithm 1, while testing M_2 is analogous, as is the case when testing more than two models. The bootstrap MJ test, in turn, can be implemented using Algorithm 2. Throughout both algorithms and the remainder of the article, the indicator function of the set S is denoted by $\mathbb{1}_S$.

Algorithm 1 Bootstrap J test

```
Initial step (original data):
Estimate models $M_1$ and $M_2$ via the maximum likelihood method, obtaining $\widehat{\boldsymbol{\beta}}_1, \widehat{\boldsymbol{\gamma}}_1$ for $M_1$ and $\widehat{\boldsymbol{\beta}}_2$ for $M_2$.
Obtain the fitted linear predictor $\widehat{\boldsymbol{\eta}}_2 = \boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2$ from $M_2$, and include it as a covariate in $M_1$.
Estimate the augmented model using the original response $\boldsymbol{y}$, and compute the test statistic $J$.
Bootstrap loop:
for $b = 1$ to $B$ do
    Generate a bootstrap sample $\boldsymbol{y}_b^*$ under $M_1$, where $y_{b,t}^* \sim \mathrm{Beta}(\widehat{\mu}_{1,t}, \widehat{\phi}_{1,t})$, with $\widehat{\mu}_{1,t} = g^{-1}(\boldsymbol{x}_{1,t}^\top \widehat{\boldsymbol{\beta}}_1)$ and $\widehat{\phi}_{1,t} = h^{-1}(\boldsymbol{z}_t^\top \widehat{\boldsymbol{\gamma}}_1)$.
    Re-estimate models $M_1$ and $M_2$ on $\boldsymbol{y}_b^*$, obtaining $\widehat{\boldsymbol{\beta}}_{1,b}^*, \widehat{\boldsymbol{\gamma}}_{1,b}^*$ and $\widehat{\boldsymbol{\beta}}_{2,b}^*$.
    Compute the fitted linear predictor $\widehat{\boldsymbol{\eta}}_{2,b}^* = \boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_{2,b}^*$.
    Estimate the augmented model using $\boldsymbol{y}_b^*$ as the response and $\widehat{\boldsymbol{\eta}}_{2,b}^*$ as an additional covariate.
    Compute the bootstrap test statistic $J_b^*$.
end for
Compute the bootstrap p-value as $p^*(J) = (1/B)\sum_{b=1}^{B} \mathbb{1}_{\{J_b^* > J\}}$.
Reject model $M_1$ if $p^*(J) < \alpha$, where $\alpha$ is the nominal significance level.
```

Algorithm 2 Bootstrap MJ test

```
Initial step (original data):
for $m = 1$ to $2$ do
    Estimate model $M_m$ via the maximum likelihood method and compute the test statistic $J_m$.
    Let $\widehat{\boldsymbol{\beta}}_m$ and $\widehat{\boldsymbol{\gamma}}_m$ denote the estimated mean and precision parameter vectors.
end for
Set $\mathrm{MJ} = \min\{J_1, J_2\}$.
Bootstrap loop:
for $b = 1$ to $B$ do
    for $m = 1$ to $2$ do
        Generate a bootstrap sample $\boldsymbol{y}_{b,m}^*$ under model $M_m$, where $y_{b,m,t}^* \sim \mathrm{Beta}(\widehat{\mu}_{m,t}, \widehat{\phi}_{m,t})$, with $\widehat{\mu}_{m,t} = g^{-1}(\boldsymbol{x}_{m,t}^\top \widehat{\boldsymbol{\beta}}_m)$ and $\widehat{\phi}_{m,t} = h^{-1}(\boldsymbol{z}_t^\top \widehat{\boldsymbol{\gamma}}_m)$.
        Re-estimate models $M_1$ and $M_2$ using $\boldsymbol{y}_{b,m}^*$ and compute $J_{b,1}^*$ and $J_{b,2}^*$.
    end for
    Set $\mathrm{MJ}_b^* = \min\{J_{b,1}^*, J_{b,2}^*\}$.
end for
Compute the bootstrap p-value as $p^*(\mathrm{MJ}) = (1/B)\sum_{b=1}^{B} \mathbb{1}_{\{\mathrm{MJ}_b^* > \mathrm{MJ}\}}$.
Reject the null hypothesis —that the true model is among the candidates tested— if $p^*(\mathrm{MJ}) < \alpha$, where $\alpha$ is the nominal significance level.
```

4. FAST DOUBLE BOOTSTRAP J AND MJ TESTS

Although the bootstrap versions of the J and MJ tests generally perform well, they can still exhibit size distortions in certain scenarios, as shown in the Monte Carlo studies presented in [4, 7]. To address this issue, an FDB scheme was proposed in [18] for using with the J test in linear regression models, demonstrating its superior performance relative to the standard bootstrap method.

The FDB method, introduced in [17], offers a computationally efficient alternative to the conventional double bootstrap. While the standard approach requires generating C second-level pseudo-samples for each first-level pseudo-sample, the FDB method reduces this burden by drawing only one second-level bootstrap sample per first-level sample.

The FDB method proceeds as follows. Under the null hypothesis, generate B first-level bootstrap samples and compute the test statistic for each pseudo-sample; that is, generate $\boldsymbol{y}_b^*$ and compute $\widehat{\tau}_b^* = \tau(\boldsymbol{y}_b^*)$, for $b = 1, \dots, B$. Then, for each $\boldsymbol{y}_b^*$, generate a single second-level bootstrap sample $\boldsymbol{y}_b^{**}$ (imposing the null hypothesis), and compute $\widehat{\tau}_b^{**} = \tau(\boldsymbol{y}_b^{**})$. The FDB p-value is given by $p^{**}(\tau) = (1/B)\sum_{b=1}^{B} \mathbb{1}_{\{\widehat{\tau}_b^* > \widehat{Q}_B^{**}(1-p^*(\tau))\}}$, where $p^*(\tau) = (1/B)\sum_{b=1}^{B} \mathbb{1}_{\{\tau_b^* > \widehat{\tau}\}}$ is the first-level bootstrap p-value, and $\widehat{Q}_B^{**}(1 - p^*(\tau))$ is the $(1-p^*(\tau))$ quantile of the second-level test statistics $\widehat{\tau}_1^{**}, \dots, \widehat{\tau}_B^{**}$. Under suitable regularity conditions, in [22], it was showed that the FDB p-value is asymptotically equivalent to that of the standard double bootstrap.

Building on the model framework above stated M_i: $g(\boldsymbol{\mu}_i) = \boldsymbol{\eta}_i = \boldsymbol{X}_i\boldsymbol{\beta}_i$ and $h(\boldsymbol{\phi}_i) = \boldsymbol{\delta}_i = \boldsymbol{Z}\boldsymbol{\gamma}_i$, we propose FDB variants of the J and MJ tests, outlined in Algorithms 3 and 4, respectively. These procedures can be readily adapted to accommodate alternative model specifications, including changes in the set of regressors or in the choice of link functions.

Algorithm 3 Fast double bootstrap J test

```
Initial step (original data):
Estimate models $M_1$ and $M_2$ via the maximum likelihood method, obtaining $\widehat{\boldsymbol{\beta}}_1, \widehat{\boldsymbol{\gamma}}_1$ for $M_1$, and $\widehat{\boldsymbol{\beta}}_2$ for $M_2$.
Compute the fitted predictor $\widehat{\boldsymbol{\eta}}_2 = \boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2$, and include it as a covariate in $M_1$.
Estimate the augmented model using the original response $\boldsymbol{y}$, and compute the test statistic $J$.
Bootstrap loop:
for $b = 1$ to $B$ do
    (i) First-level sample:
    Generate $\boldsymbol{y}_b^*$ under model $M_1$, where $y_{b,t}^* \sim \mathrm{Beta}(\widehat{\mu}_{1,t}, \widehat{\phi}_{1,t})$, where $\widehat{\mu}_{1,t} = g^{-1}(\boldsymbol{x}_{1,t}^\top \widehat{\boldsymbol{\beta}}_1)$, $\widehat{\phi}_{1,t} = h^{-1}(\boldsymbol{z}_t^\top \widehat{\gamma}_1)$.
    Re-estimate models $M_1$ and $M_2$ using $\boldsymbol{y}_b^*$, obtaining $\widehat{\boldsymbol{\beta}}_{1,b}^*, \widehat{\boldsymbol{\gamma}}_{1,b}^*, \widehat{\boldsymbol{\beta}}_{2,b}^*$.
    Compute $\widehat{\boldsymbol{\eta}}_{2,b}^* = \boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_{2,b}^*$.
    Estimate the augmented model using $\boldsymbol{y}_b^*$ as the response and $\widehat{\eta}_{2,b}^*$ as an additional covariate and compute the bootstrap test statistic $J_b^*$.
    (ii) Second-level sample (FDB):
    Generate $\boldsymbol{y}_b^{**}$ under the re-estimated model $M_1$ , where $y_{b,t}^{**} \sim \mathrm{Beta}(\widehat{\mu}_{1,t}^*, \widehat{\phi}_{1,t}^*)$, where $\widehat{\mu}_{1,t}^* = g^{-1}(\boldsymbol{x}_t^\top \widehat{\boldsymbol{\beta}}_{1,b}^*)$, $\widehat{\phi}_{1,t}^* = h^{-1}(\boldsymbol{z}_t^\top \widehat{\boldsymbol{\gamma}}_{1,b}^*)$.
    Re-estimate the augmented model using $\boldsymbol{y}_b^{**}$, and compute the second-level test statistic $J_b^{**}$.
end for
Compute the first-level bootstrap p-value as $p^*(J) = (1/B)\sum_{b=1}^{B} \mathbb{1}_{\{J_b^* > J\}}$.
Obtain the $(1 - p^*(J))$ quantile of $\{J_1^{**}, \dots, J_B^{**}\}$, denoted $\widehat{Q}_B^{**}(1 - p^*(J))$.
Calculate the FDB p-value as $p_F^{**}(J) = (1/B)\sum_{b=1}^{B} \mathbb{1}_{\{J_b^* > \widehat{Q}_B^{**}(1-p^*(J))\}}$.
Reject model $M_1$ if $p_F^{**}(J) < \alpha$, where $\alpha$ denotes the nominal significance level.
```

Algorithm 4 Fast double bootstrap MJ test

```
Initial step (original data):
for $m = 1$ to $2$ do
    Estimate model $M_m$ via the maximum likelihood method, obtaining $\widehat{\boldsymbol{\beta}}_m$ and $\widehat{\boldsymbol{\gamma}}_m$, which are the maximum likelihood estimates of the mean and precision parameters.
    Compute the test statistic $J_m$.
end for
Set $\mathrm{MJ} = \min\{J_1, J_2\}$.
Bootstrap loop:
for $b = 1$ to $B$ do
    (i) First-level samples:
    for $m = 1$ to $2$ do
        Generate $\boldsymbol{y}^*_{b,m}$ under model $M_m$, where $y^*_{b,m,t} \sim \mathrm{Beta}(\widehat{\mu}_{m,t}, \widehat{\phi}_{m,t})$, with $\widehat{\mu}_{m,t} = g^{-1}(\boldsymbol{x}^\top_{m,t}\widehat{\boldsymbol{\beta}}_m)$, $\widehat{\phi}_{m,t} = h^{-1}(\boldsymbol{z}^\top_t\widehat{\boldsymbol{\gamma}}_m)$.
        Re-estimate models $M_1$ and $M_2$ using $\boldsymbol{y}^*_{b,m}$ and compute $J^*_{b,1}$ and $J^*_{b,2}$.
    end for
    Set $\mathrm{MJ}^*_b = \min\{J^*_{b,1}, J^*_{b,2}\}$.
    (ii) Second-level samples (FDB):
    for $m = 1$ to $2$ do
        Generate $\boldsymbol{y}^{**}_{b,m}$ under the re-estimated model $M_m$, where $y^{**}_{b,m,t} \sim \mathrm{Beta}(\widehat{\mu}^*_{m,t}, \widehat{\phi}^*_{m,t})$, with $\widehat{\mu}^*_{m,t} = g^{-1}(\boldsymbol{x}^\top_t\widehat{\boldsymbol{\beta}}^*_{m,b})$, $\widehat{\phi}^*_{m,t} = h^{-1}(\boldsymbol{z}^\top_t\widehat{\boldsymbol{\gamma}}^*_{m,b})$.
        Re-estimate models $M_1$ and $M_2$ on $\boldsymbol{y}^{**}_{b,m}$ and compute $J^{**}_{b,1}$ and $J^{**}_{b,2}$.
    end for
    Set $\mathrm{MJ}^{**}_b = \min\{J^{**}_{b,1}, J^{**}_{b,2}\}$.
end for
Compute the first-level bootstrap p-value as $p^*(\mathrm{MJ}) = (1/B)\sum_{b=1}^{B} \mathbb{1}_{\{\mathrm{MJ}^*_b > \mathrm{MJ}\}}$.
Obtain the $(1 - p^*(\mathrm{MJ}))$ quantile of $\{\mathrm{MJ}^{**}_1, \ldots, \mathrm{MJ}^{**}_B\}$, denoted $\widehat{Q}^{**}_B(1 - p^*(\mathrm{MJ}))$.
Calculate the FDB p-value as $p^{**}_F(\mathrm{MJ}) = (1/B)\sum_{b=1}^{B} \mathbb{1}_{\{\mathrm{MJ}^*_b > \widehat{Q}^{**}_B(1-p^*(\mathrm{MJ}))\}}$.
Reject the null hypothesis (that the true model is among those tested) if $p^{**}_F(\mathrm{MJ}) < \alpha$, where $\alpha$ denotes the nominal significance level.
```

5. SIMULATION EVIDENCE

This section presents Monte Carlo simulation results evaluating the finite-sample performance of the J and MJ tests, along with their bootstrap-based versions, in small samples. The simulations are based on 10,000 Monte Carlo replications.

Tests were conducted at nominal significance levels $\alpha = 10\%, 5\%$, and 1%, using 1,000 bootstrap replications. Covariate values were generated as independent draws from the standard uniform distribution. We consider three sample sizes, $n = 20, 30, 40$, with $\mu \in (0.29, 0.95)$ and a constant degree of heterogeneity across all designs, measured by the ratio between the maximum and minimum values of the precision parameter ϕ_t.

Six different scenarios are examined, in which the competing models differ either in their set of regressors or in the link functions adopted for the mean and/or precision submodels. All simulations were conducted using the matrix programming language Ox [23]. The numerical maximizations of the beta regression log-likelihood function were carried out using the BFGS nonlinear optimization algorithm with analytical first derivatives. The starting values for the parameters $\beta_1, \ldots, \beta_k$ were chosen following the approach proposed in [5], originally developed for beta regression models with fixed precision. The initial value for γ_1 was obtained by applying the link function h to the fixed precision starting value suggested by those authors, and the remaining precision parameters, $\gamma_2, \ldots, \gamma_m$, were initialized at zero.

At the outset, we consider a scenario in which two competing models differ in the regressors of the mean submodel. The models are specified as

$$M_1\colon \log\left(\frac{\mu_t}{1-\mu_t}\right) = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3}, \quad M_2\colon \log\left(\frac{\mu_t}{1-\mu_t}\right) = \beta_1 + \beta_2 x_{t2} + \beta_4 x_{t4},$$
$$\log(\phi_t) = \gamma_1 + \gamma_2 x_{t2}, \qquad \log(\phi_t) = \gamma_1 + \gamma_2 x_{t2}.$$

Table 1 reports the null rejection rates of the tests. The results indicate that, for all tests, rejection rates tend to converge to the nominal significance levels as the sample size increases. The bootstrap-based versions of the J and MJ tests consistently outperform their asymptotic counterparts, which tend to over-reject.

For example, when $n = 20$ and $\alpha = 5\%$, the null rejection rate of the asymptotic MJ test is 8.66%, while the rates for the bootstrap MJ* and the FDB MJ** tests are 4.53% and 4.59%, respectively. Notably, in most cases, the FDB test exhibits better control of Type I error than the standard bootstrap test, yielding rejection rates closer to the nominal level.

Table 1 Null rejection rates (%) when the models differ in the mean submodel regressors.

α	10%			5%			1%		
n	20	30	40	20	30	40	20	30	40
J	14.91	12.29	11.05	8.72	6.59	5.99	2.78	1.86	1.55
J^*	8.62	8.44	8.73	4.46	4.29	4.22	0.93	0.94	1.07
J^{**}	8.74	8.61	8.74	4.50	4.42	4.26	0.94	1.06	1.11
MJ	14.88	12.29	11.05	8.66	6.58	5.99	2.73	1.86	1.55
MJ*	8.73	8.47	8.74	4.53	4.33	4.22	0.97	0.94	1.07
MJ**	8.75	8.61	8.72	4.59	4.40	4.24	0.97	1.05	1.11

The second scenario considers models that differ in the regressors of the precision submodel. The competing models are

$$M_1\colon \log\left(\frac{\mu_t}{1-\mu_t}\right) = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3}, \quad M_2\colon \log\left(\frac{\mu_t}{1-\mu_t}\right) = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3},$$
$$\log(\phi_t) = \gamma_1 + \gamma_2 x_{t2}, \qquad \log(\phi_t) = \gamma_1 + \gamma_3 x_{t3}.$$

The results in Table 2 indicate that the J^* and J^{**} tests perform similarly, with the standard bootstrap version J^* yielding rejection rates slightly closer to the nominal significance levels in most cases. In contrast, MJ** consistently outperforms MJ*. For example, when $n = 20$ and $\alpha = 10\%$, the null rejection rates of the MJ** and MJ* tests are 8.38% and 7.22%, respectively.

Table 2 Null rejection rates (%) when the models differ in the precision submodel regressors.

α	10%			5%			1%		
n	20	30	40	20	30	40	20	30	40
J	17.79	13.80	12.63	10.87	7.45	7.18	3.63	2.04	1.57
J^*	11.38	10.18	10.34	6.08	5.08	5.45	1.45	1.10	1.06
J^{**}	11.39	10.14	10.34	6.26	5.06	5.53	1.53	1.16	1.09
MJ	6.26	5.37	5.48	3.00	2.31	2.53	0.59	0.42	0.37
MJ*	7.22	7.87	8.52	3.26	3.20	3.86	0.52	0.39	0.58
MJ**	8.38	8.53	8.86	3.76	3.53	3.96	0.62	0.58	0.71

In the third scenario, we consider two models that differ in the regressors of both the mean and precision submodels. The competing models are

$$M_1\colon \log\left(\frac{\mu_t}{1-\mu_t}\right) = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3}, \quad M_2\colon \log\left(\frac{\mu_t}{1-\mu_t}\right) = \beta_1 + \beta_2 x_{t2} + \beta_4 x_{t4},$$

$$\log(\phi_t) = \gamma_1 + \gamma_2 x_{t2}, \qquad \log(\phi_t) = \gamma_1 + \gamma_3 x_{t3}.$$

The results reported in Table 3 indicate that, except when $n = 20$ and $\alpha = 1\%$, the FDB versions of the J and MJ tests outperform their standard bootstrap counterparts. For example, when $n = 20$ and $\alpha = 5\%$, the null rejection rates of the J^* and MJ* tests are 4.75% and 4.85%, respectively, compared to 4.93% and 4.97% for the J^{**} and MJ** tests.

Table 3 Null rejection rates (%) when the models differ in the mean and precision submodels regressors.

α	10%			5%			1%		
n	20	30	40	20	30	40	20	30	40
J	20.16	14.86	13.20	12.96	8.51	7.39	4.69	2.26	1.97
J^*	9.58	9.12	9.38	4.75	4.35	4.79	1.28	0.91	0.93
J^{**}	9.83	9.23	9.54	4.93	4.40	4.96	1.37	1.00	1.04
MJ	20.10	14.85	13.20	12.89	8.49	7.39	4.61	2.25	1.97
MJ*	9.59	9.14	9.38	4.85	4.35	4.81	1.25	0.92	0.93
MJ**	9.85	9.21	9.55	4.97	4.41	4.98	1.33	1.00	1.02

For the fourth scenario, we consider the case in which the two competing models differ in their mean link functions. The models are given by

$$M_1\colon -\log(-\log \mu_t) = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3}, \quad M_2\colon \log\left(\frac{\mu_t}{1-\mu_t}\right) = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3},$$

$$\log(\phi_t) = \gamma_1 + \gamma_2 x_{t2}, \qquad \log(\phi_t) = \gamma_1 + \gamma_2 x_{t2}.$$

Table 4 reports the null rejection rates of the tests. The results indicate that the MJ test outperforms the J test, and that the bootstrap-based versions provide improved performance over their asymptotic counterparts. Notably, the FDB tests generally outperform the standard bootstrap tests. For example, when $n = 20$ and $\alpha = 10\%$, the null rejection rate of the MJ* test is 10.32%, whereas that of the MJ** test is 10.04%.

Table 4 Null rejection rates (%) when the models differ in the mean submodel link function.

α	10%			5%			1%		
n	20	30	40	20	30	40	20	30	40
J	16.15	13.89	12.47	9.67	7.86	6.74	3.01	1.98	1.59
J^*	9.95	10.15	9.98	5.26	4.90	4.87	1.10	1.19	1.15
J^{**}	9.93	10.24	10.03	5.20	4.98	5.04	1.08	1.21	1.17
MJ	11.55	9.31	8.60	6.64	4.92	4.30	1.93	1.18	0.99
MJ*	10.32	10.21	10.60	5.42	5.02	5.21	1.07	1.07	1.07
MJ**	10.04	10.00	10.46	5.27	5.00	5.15	1.00	1.14	1.24

We now consider the case in which the two competing models differ in the precision link function. The models are specified as

$$M_1\colon\ \log\left(\frac{\mu_t}{1-\mu_t}\right) = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3}, \quad M_2\colon\ \log\left(\frac{\mu_t}{1-\mu_t}\right) = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3},$$
$$\sqrt{\phi_t} = \gamma_1 + \gamma_2 x_{t2}, \qquad \log(\phi_t) = \gamma_1 + \gamma_2 x_{t2}.$$

The results in Table 5 indicate that the J^* test generally outperforms J^{**}, whereas for the MJ tests, the FDB version MJ** tends to perform better than the standard bootstrap version MJ*. For instance, when $n = 30$ and $\alpha = 10\%$, the null rejection rates of MJ* and MJ** are 10.78% and 10.34%, respectively.

Table 5 Null rejection rates (%) when the models differ in the precision submodel link function.

α		10%			5%			1%	
n	20	30	40	20	30	40	20	30	40
J	19.18	15.06	13.12	11.92	8.45	7.28	3.70	2.22	1.65
J^*	10.76	10.25	9.88	5.38	5.29	5.12	1.06	1.04	0.98
J^{**}	10.80	10.15	9.92	5.54	5.34	5.18	1.10	1.11	1.02
MJ	16.84	12.96	11.45	10.14	6.18	2.58	2.90	1.76	1.32
MJ*	11.38	10.78	10.25	5.88	5.41	5.24	1.22	1.13	1.05
MJ**	10.88	10.34	10.12	5.60	5.34	5.15	1.32	1.19	1.08

Lastly, we examine the scenario in which the two models differ in both the mean and precision link functions. The competing models are stated as

$$M_1\colon\ -\log(-\log(\mu_t)) = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3}, \quad M_2\colon\ \log\left(\frac{\mu_t}{1-\mu_t}\right) = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3},$$
$$\sqrt{\phi_t} = \gamma_1 + \gamma_2 x_{t2}, \qquad \log(\phi_t) = \gamma_1 + \gamma_2 x_{t2}.$$

The results in Table 6 show that, in most cases, the J^* test outperforms J^{**}, whereas MJ** performs better than MJ*. For example, when $n = 20$ and $\alpha = 10\%$, the null rejection rates of J^*, J^{**}, MJ*, and MJ** are 10.19%, 10.23%, 10.46%, and 10.15%, respectively.

Table 6 Null rejection rates (%) when the models differ in the mean and precision submodels link functions.

α		10%			5%			1%	
n	20	30	40	20	30	40	20	30	40
J	21.10	15.65	13.72	13.22	8.99	7.79	4.53	2.70	2.00
J^*	10.19	9.59	9.89	5.30	5.10	4.88	1.11	1.18	1.06
J^{**}	10.23	9.77	9.75	5.47	5.16	5.05	1.18	1.32	1.04
MJ	16.82	12.21	10.40	9.96	6.85	5.95	3.12	1.89	1.33
MJ*	10.46	10.20	9.98	5.45	5.06	5.24	1.12	1.20	1.11
MJ**	10.15	9.89	9.89	5.40	5.08	5.06	1.12	1.19	1.09

Overall, the Monte Carlo results show that the J and MJ tests based on the FDB exhibit substantially smaller size distortions than their non-bootstrap counterparts across the scenarios considered. While the improvements over the standard bootstrap versions of the tests are generally modest, they are not negligible. These additional gains in accuracy come at a very low computational cost, as the FDB procedure requires only a single replication at the second level of resampling. This computational efficiency makes the FDB approach particularly attractive in practice, especially in situations where the bootstrap p-value lies close to the nominal significance level. In such cases, the improved accuracy of FDB-based inference may influence empirical conclusions and enhance the reliability of hypothesis testing in small samples.

6. EMPIRICAL APPLICATION

In this section, we present an empirical application aimed at modeling the proportion of income spent on food as a function of household income and the number of people in the household. Parameter estimates for the beta regression models are obtained by numerically maximizing the corresponding log-likelihood functions using the BFGS optimization algorithm with analytical first derivatives, implemented in the Ox programming language. The starting values for the model parameters were chosen as described in the previous section. Specifically, the approach proposed in [5] was used to initialize the mean submodel parameters, while the initial value for the intercept in the precision submodel was obtained by applying the link function h to the fixed-precision starting value suggested by those authors, with the remaining precision parameters set to zero.

The R computing environment [24] was used exclusively to perform model specification tests (RESET tests) and to generate residual empirical quantile versus theoretical quantile (QQ) plots with simulated envelopes. These plots are based on the standardized weighted residual 2 proposed in [25], using 500 replications. All RESET tests were carried out at the 10% significance level. In R, parameter estimates were obtained using the *betareg* package (CRAN.R-project.org/package=betareg), which provides maximum likelihood estimation for beta regression models with fixed or varying precision.

We use the household expenditure data presented in [26] (Table 15.4), which include information on food expenditure, income (X_2), and the number of people in the household (X_3) from a U.S. city. In [5], the proportion of income spent on food (Y) was modeled as a function of X_2 and X_3 using a beta regression model with constant precision. Subsequently, a varying-precision model was proposed in [6], adopting a logit link for the mean submodel and a logarithmic link for the precision submodel, specified as

$$\log\left(\frac{\mu_t}{1-\mu_t}\right) = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3}, \quad \log(\phi_t) = \gamma_1 + \gamma_2 x_{t2} + \gamma_3 x_{t3}, \quad t = 1, \ldots, 38,$$

where x_1, x_2, x_3 are the observed values of X_1, X_2, X_3, respectively.

After performing a series of nested hypothesis tests (not reported here) using the FDB version of the LR test [12], we adopted the initial model formulated as

$$M_1\colon \log\left(\frac{\mu_t}{1-\mu_t}\right) = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3}, \quad \log(\phi_t) = \gamma_1 + \gamma_2 x_{t3}.$$

To investigate potential signs of model misspecification, we applied the procedure proposed in [27], based on the RESET test introduced in [28]. The RESET test (statistic = 5.024, p-value = 0.025) indicates evidence of misspecification. Accordingly, we assessed whether alternative link functions could improve model fit by testing M_1 against models with different link specifications for the mean submodel.

Table 7 presents the results of these comparisons. The findings indicate that all tested models exhibit some degree of misspecification. At the 10% nominal significance level, all models would be rejected. At the 5% level, only the models with logit and Cauchy links remain plausible. However, direct comparisons show that the logit model is rejected at the 5% level by both the asymptotic (p-value = 0.018) and standard bootstrap (p-value = 0.018) versions of the J test, and at the 1% level by the FDB version (p-value = 0.009). Although the Cauchy link model is not rejected by the standard bootstrap J test at the 5% level (p-value = 0.061), it is rejected by both the asymptotic and FDB versions (p-value = 0.042). The MJ test reinforces these conclusions, with the FDB version providing stronger evidence against the null hypothesis that the true model belongs to the tested class (p-value = 0.031).

Table 7 p-values from the J and MJ tests —along with their bootstrap and FDB variants— for assessing alternative link functions in food expenditure models.

Models	J_1	J_1^*	J_1^{**}	J_2	J_2^*	J_2^{**}	MJ	MJ*	MJ**
logit versus log-log	0.055	0.081	0.077	0.029	0.039	0.045	0.055	0.072	0.071
logit versus Cauchy	0.018	0.018	0.009	0.042	0.061	0.042	0.042	0.044	0.031
logit versus clog-log	0.023	0.033	0.033	0.030	0.046	0.046	0.030	0.042	0.042
log-log versus Cauchy	0.016	0.026	0.023	0.064	0.075	0.056	0.064	0.043	0.029
log-log versus clog-log	0.022	0.041	0.041	0.052	0.083	0.096	0.052	0.066	0.070
clog-log versus Cauchy	0.022	0.033	0.035	0.039	0.064	0.078	0.039	0.047	0.069

For comparison, we define the Cauchy link model stated as

$$M_2\colon\ \tan(\pi(\mu_t - 0.5)) = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3}, \quad \log(\phi_t) = \gamma_1 + \gamma_2 x_{t3}.$$

To evaluate the model fit, we considered several metrics, including the pseudo-R^2_{LR} [29] and its adaptation for variable precision beta regression models, $\bar{R}^2_{\text{LR}}$ [30]. Additionally, we used the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). Table 8 presents these goodness-of-fit measures for both M_1 and M_2, highlighting the superior fit of M_2.

Table 8 Goodness-of-fit measures, food expenditure data.

Model	R^2_{LR}	$\bar{R}^2_{\text{LR}}$	AIC	BIC
M_1	0.517	0.452	−88.370	−80.182
M_2	0.538	0.475	−90.022	−81.834

However, the RESET test for M_2 (statistic = 4.193, p-value = 0.041) also indicates possible model misspecification, which may explain the rejections observed in the J^{**} and MJ** tests. Figure 1 presents normal probability plots for both models, showing no strong evidence against either model, as the residuals remain within the simulated confidence envelopes.

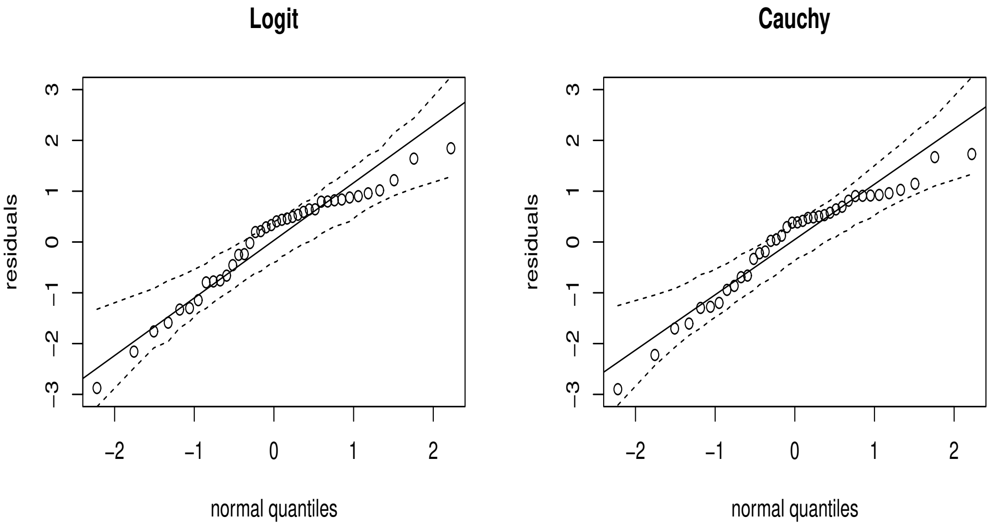

Figure 1 Residual QQ plot with simulated envelopes for Models M_1 (left panel) and M_2 (right panel), food expenditure data.

Although the model with the Cauchy link appears to provide a better fit, further refinement is warranted. Accordingly, we considered a more flexible specification that incorporates the interaction between covariates x_{t2} and x_{t3} in both the mean and precision submodels, given by

$$\begin{aligned} \tan\big(\pi(\mu_t - 0.5)\big) &= \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + \beta_4 x_{t2} x_{t3}, \\ \log(\phi_t) &= \gamma_1 + \gamma_2 x_{t2} + \gamma_3 x_{t3} + \gamma_4 x_{t2} x_{t3}. \end{aligned}$$

We then carried out a sequence of nested hypothesis tests (not reported here), using the FDB version of the LR test to guide model selection. Based on the results, we adopted the reduced model formulated as

$$M_3\colon\ \tan\big(\pi(\mu_t - 0.5)\big) = \beta_1 + \beta_2 x_{t3} + \beta_3 x_{t2} x_{t3}, \quad \log(\phi_t) = \gamma_1 + \gamma_2 x_{t3}.$$

Note that models M_2 and M_3 are non-nested with respect to the regressors included in the mean submodel. The p-values reported in Table 9 indicate that none of the versions of the J test reject either model at the 5% significance level. Similarly, the MJ test —whether in its standard, bootstrap, or FDB form— fails to reject the null hypothesis that one of the two models is correctly specified.

Since $J_3 < J_2$ ($J_2 = 3.380$ and $J_3 = 1.392$), model M_3 is selected. Additionally, the RESET test applied to M_3 (statistic = 0.543, p-value = 0.461) does not indicate model misspecification at conventional significance levels.

Table 9 p-values from the J and MJ tests —along with their bootstrap and FDB versions— for comparing models M_2 and M_3.

Models	J_2	J_2^*	J_2^{**}	J_3	J_3^*	J_3^{**}	MJ	MJ*	MJ**
M_2 and M_3	0.066	0.092	0.103	0.238	0.286	0.283	0.238	0.198	0.156

Figure 2 presents the residual QQ plot with simulated envelopes for model M_3. All residuals lie within the simulated confidence bands, providing no evidence of model misspecification or outlying observations. Furthermore, the goodness-of-fit statistics in Table 10 confirm that model M_3 achieves a better fit than the competing alternatives previously reported in Table 8.

Table 11 presents the parameter estimates and corresponding standard errors for model M_3. In addition, the variance inflation factor for this model is 2.522, indicating that multicollinearity among the regressors in the mean submodel is not a concern.

Table 10 Goodness-of-fit measures for model M_3, food expenditure data.

Model	R^2_{LR}	$\bar{R}^2_{\text{LR}}$	AIC	BIC
M_3	0.558	0.498	−91.279	−83.541

Table 11 Parameter estimates and standard errors for model M_3, food expenditure data.

	β_1	β_2	β_3	γ_1	γ_2
Estimate	−1.231	−0.305	−0.003	5.022	−0.324
Standard error	0.123	0.053	0.001	0.535	0.134

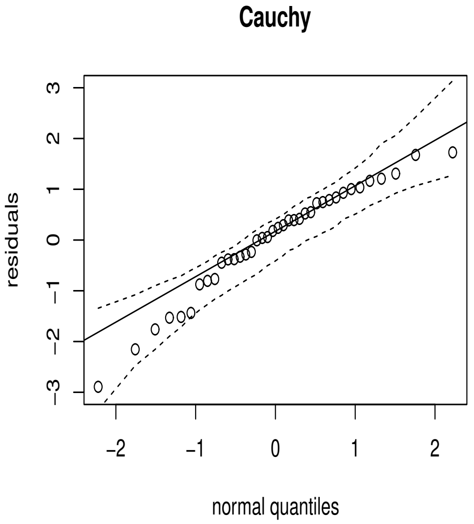

Figure 2 Residual QQ plot with simulated envelopes for model M_3, food expenditure data.

While the main purpose of the empirical application is to illustrate the use of nonnested hypothesis testing and the advantages of the FDB method, the final model (M_3) also reveals interesting patterns in household food expenditure behavior. In this model, the conditional mean of the response —that is, the proportion of income spent on food— is linked to covariates via a Cauchy link function and includes an interaction term between income (X_2) and household size (X_3). Due to the nonlinear form of the link function, the marginal effects of these covariates on μ_t are not constant and depend on the values of both income and household size. This interaction implies that the influence of income on the proportion of income spent on food is modulated by household size: for example, the effect of increasing income may be stronger or weaker depending on whether the household is small or large. The precision submodel also indicates that variability in food expenditure shares is associated with household size. Together, these results suggest the presence of scale effects and heterogeneous consumption behavior across different household profiles.

Based on the selected model, we can estimate the marginal effect of income on the mean proportion of income spent on food. This marginal effect is $\partial\mu_t/\partial x_{t2} = \beta_3 x_{t3}/\left(\pi(1+\eta_t^2)\right)$, where $\eta_t = \tan\bigl(\pi(\mu_t - 0.5)\bigr)$ is the linear predictor in the mean submodel. This quantity is estimated by replacing the parameters with their maximum likelihood estimates. Figure 3 displays the estimated marginal effects of income as a function of income itself, for three different household sizes: 1, 4, and 7. The estimated effects are negative across all income levels and household sizes, indicating that increases in income consistently lead to a reduction in the expected proportion of income allocated to food.

The marginal effect is closer to zero (that is, less negative) for smaller households. The curve corresponding to single-person households lies above the others, suggesting that increases in income have a relatively smaller impact on food expenditure in these cases. In contrast, the curve for households with seven members lies well below the others, revealing that the effect of income is considerably stronger (more negative) in larger households. Moreover, household size (X_3) also influences the shape of the marginal effect curve. For single-person households, the curve is nearly flat, indicating that the marginal effect varies little with income. However, for household sizes 4 and 7, the curves exhibit a clear upward trend (toward zero), suggesting that the magnitude of the negative effect decreases as income increases. This pattern implies that, for larger households, increases in income have a stronger impact on reducing the share of income spent on food, particularly at lower income levels.

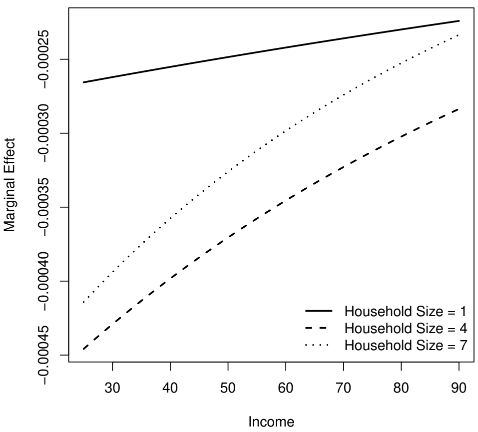

Figure 3 Estimated impact of income on the mean proportion of income spent on food.

7. Conclusions

In regression analysis, once the candidate models have been estimated, it is standard practice to conduct a series of tests to assess their compatibility with the observed data. For non-nested models —where one specification cannot be derived from another through linear parameter restrictions— the J and MJ tests are widely regarded as effective tools. These procedures are fundamental for model comparison, offering a rigorous framework for selecting among competing non-nested alternatives. Specifically, in the context of beta regression models, the J and MJ tests were adapted [7] to accommodate the peculiarities of this class of models. Their work also introduced bootstrap-based versions of the tests, providing a more robust approach in the presence of heteroskedasticity and non-normality, which are common challenges in beta regression settings.

This study contributes to this line of research by proposing fast double bootstrap versions of the J and MJ tests specifically tailored for beta regression models. The main motivation behind this refinement is to enhance the precision and reliability of these tests, particularly in small samples where conventional asymptotic methods often perform poorly.

We evaluated the performance of the fast double bootstrap tests through extensive Monte Carlo simulations designed to assess their effectiveness in non-nested hypothesis testing. The results were compelling: compared to the standard bootstrap versions, the fast double bootstrap test consistently produced test sizes closer to nominal significance levels. Both the J and MJ tests suffer from substantial size distortions in small samples, and our findings reinforce the recommendation to adopt bootstrap-based corrections in such contexts. Although the improvements offered by the fast double bootstrap test over the standard bootstrap are generally modest, they come at a minimal computational cost, as only a single second-level resample is required per bootstrap replication. These marginal gains in accuracy can be particularly valuable when the bootstrap p-value lies near the rejection threshold, potentially influencing empirical conclusions. For this reason, we recommend the use of the fast double bootstrap test as a practical and effective refinement of bootstrap-based inference.

We also presented an empirical application focused on modeling the proportion of income allocated to food expenditure as a function of household income and household size. In this context, we identified a model that provided a superior fit relative to commonly used specifications. This conclusion was supported by non-nested hypothesis testing conducted with the fast double bootstrap test. The application effectively highlighted the practical value of the proposed methodology for empirical model selection.

In summary, the contributions of this study lie in both the methodological developments and the practical insights it offers. By enhancing the J and MJ tests with fast double bootstrap tests, this work improves the accuracy of model comparisons in beta regression, providing a more reliable and accessible tool for researchers and practitioners. These advancements not only deepen our understanding of model testing but also help ensure that the selected model better reflects the underlying data structure.

There are several promising directions for future research. For instance, it would be worthwhile to investigate the behavior of the J and MJ tests with bootstrap resampling in extended versions of the beta regression model, such as inflated, mixed, and dynamic models; see, respectively, [31, 32, 33]. Another avenue is the exploration of alternative model classes for doubly bounded response variables, such as the simplex regression model [34].

STATEMENTS

Acknowledgements

The authors thank two anonymous reviewers for helpful comments and suggestions that improved the quality of the article.

Author contributions

Conceptualization: F. Cribari-Neto; data curation: F.P. Lima; formal analysis: F.P. Lima, F. Cribari-Neto; investigation: F.P. Lima, F. Cribari-Neto; methodology: F.P. Lima, F. Cribari-Neto; writing original draft: F.P. Lima, F. Cribari-Neto; writing review and editing: F.P. Lima, F. Cribari-Neto. All authors have read and agreed to the published version of the article.

Conflicts of interest

The authors declare no conflict of interest.

Data and code availability

The data used in our empirical analysis, along with computer code for conducting nonnested hypothesis testing (using asymptotic, bootstrap, and fast double bootstrap procedures) in one representative case, are publicly available at github.com/cribari/nonnested-1.

Declaration on the use of artificial intelligence (AI) technologies

The authors declare that no generative AI was used in the preparation of this article.

Funding

This work was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) [grant number 304646/2023-7] from Brazil.

Open access statement

References

[1] Davidson, R. and MacKinnon, J.G., 1981. Several tests for model specification in the presence of alternative hypotheses. Econometrica, 49, 781–793. DOI: 10.2307/1911522

[2] Yanqin, F. and Qi, L., 1995. Bootstrapping J-type tests for non-nested regression models. Economics Letters, 48, 107–112. DOI: 10.1016/0165-1765(94)00595-S

[3] Godfrey, L.G., 1998. Tests of non-nested regression models some results on small sample behaviour and the bootstrap. Journal of Econometrics, 84, 59–74. DOI: 10.1016/S0304-4076(97)00079-1

[4] Hagemann, A., 2012. A simple test for regression specification with non-nested alternatives. Journal of Econometrics, 166, 247–254. DOI: 10.1016/j.jeconom.2011.09.037

[5] Ferrari, S.L.P. and Cribari-Neto, F., 2004. Beta regression for modelling rates and proportions. Journal of Applied Statistics, 31, 799–815. DOI: 10.1080/0266476042000214501

[6] Cribari-Neto, F. and Zeileis, A., 2010. Beta regression in R. Journal of Statistical Software, 34(2), 1–24. DOI: 10.18637/jss.v034.i02

[7] Cribari-Neto, F. and Lucena, S.E.F., 2015. Nonnested hypothesis testing in the class of varying dispersion beta regressions. Journal of Applied Statistics, 42, 967–985. DOI: 10.1080/02664763.2014.993368

[8] Rigby, R.A. and Stasinopoulos, D.M., 2005. Generalized additive models for location, scale and shape. Journal of the Royal Statistical Society C, 54, 507–554. DOI: 10.1111/j.1467-9876.2005.00510.x

[9] Cribari-Neto, F. and Lucena, S.E.F., 2017. Non-nested hypothesis testing inference for GAMLSS models. Journal of Statistical Computation and Simulation, 87, 1189–1205. DOI: 10.1080/00949655.2016.1255946

[10] Ouysse, R., 2011. Computationally efficient approximation for the double bootstrap mean bias correction. Economics Bulletin, 31, 2388–2403. URL: core.ac.uk/download/pdf/6266542.pdf

[11] Davidson, R. and Trokić, M., 2020. The fast iterated bootstrap. Journal of Econometrics, 218, 451–475. DOI: 10.1016/j.jeconom.2020.04.025

[12] Lima, F.P. and Cribari-Neto, F., 2020. Bootstrap-based testing inference in beta regression. Brazilian Journal of Probability and Statistics, 34, 18–34. DOI: 10.1214/18-BJPS412

[13] Lamarche, J.F., 2004. The numerical performance of fast bootstrap procedures. Computational Economics, 23, 379–389. DOI: 10.1023/B:CSEM.0000026805.91428.af

[14] Omtzigt, P. and Fachin, S., 2006. The size and power of bootstrap and Bartlett-corrected tests of hypotheses on the cointegrating vectors. Econometric Reviews, 25, 41–60. DOI: 10.1080/07474930500545439

[15] Davidson, J., 2006. Alternative bootstrap procedures for testing cointegration in fractionally integrated processes. Journal of Econometrics, 133, 741–777. DOI: 10.1016/j.jeconom.2005.06.012

[16] Ahlgren, N. and Antell, J., 2008. Bootstrap and fast double bootstrap tests of cointegration rank with financial time series. Computational Statistics and Data Analysis, 52, 4754–4767. DOI: 10.1016/j.csda.2008.03.025

[17] Davidson, R. and MacKinnon, J.G., 2007. Improving the reliability of bootstrap tests with the fast double bootstrap. Computational Statistics and Data Analysis, 51, 3259–3281. DOI: 10.1016/j.csda.2006.04.001

[18] Davidson, R. and MacKinnon, J.G., 2002. Fast double bootstrap tests of nonnested linear regression models. Econometric Reviews, 21, 419–429. DOI: 10.1081/ETC-120015384

[19] Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P., 1992. Numerical Recipes in C: The Art of Scientific Computing. Cambridge University Press, New York, NY, US.
[20] Nocedal, J. and Wright, S.J., 2006. Numerical Optimization. Springer, New York, NY, US.
[21] Fahrmeir, L. and Kaufmann, H., 1985. Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. Annals of Statistics, 13, 342–368. DOI: 10.1214/aos/1176346597
[22] MacKinnon, J.G., 2006. Applications of the fast double bootstrap. Queen's Economics Department, Working Paper. DOI: 10.22004/ag.econ.273459
[23] Doornik, J.A., 2021. An Object-oriented Matrix Language - Ox 9. Timberlake Consultants Press, Richmond, UK.
[24] R Core Team, 2025. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
[25] Espinheira, P.L., Ferrari, S.L.P. and Cribari-Neto, F., 2008. On beta regression residuals. Journal of Applied Statistics, 35, 407–419. DOI: 10.1080/02664760701834931
[26] Griffiths, W., Hill, R. and Judge, G., 1993. Learning and Practicing Econometrics. Wiley, New York, NY, US.
[27] Pereira, T.L. and Cribari-Neto, F., 2014. Detecting model misspecification in inflated beta regressions. Communications in Statistics: Simulation and Computation, 43, 631–656. DOI: 10.1080/03610918.2012.712183
[28] Ramsey, J.B., 1969. Tests for specification errors in classical linear least-squares regression analysis. Journal of the Royal Statistical Society B, 31, 350–371. DOI: 10.1111/j.2517-6161.1969.tb00796.x
[29] Nagelkerke, N.J.D., 1991. A note on a general definition of the coefficient of determination. Biometrika, 78, 691–692. DOI: 10.1093/biomet/78.3.691
[30] Bayer, F.M. and Cribari-Neto, F., 2017. Model selection criteria in beta regression with varying dispersion. Communications in Statistics: Simulation and Computation, 46, 729–746. DOI: 10.1080/03610918.2014.977918
[31] Ospina, R. and Ferrari, S.L.P., 2012. A general class of zero-or-one inflated beta regression models. Computational Statistics and Data Analysis, 56, 1609–1623. DOI: 10.1016/j.csda.2011.10.005
[32] Bonat, W.H., Ribeiro Jr, P.J., and Shimakura, S.E., 2015. Bayesian analysis for a class of beta mixed models. Chilean Journal of Statistics, 6, 3–13. URL: soche.cl/chjs/volumes/06/01/Bonat_etal(2015).pdf
[33] Scher, V.T., Cribari-Neto, F., and Bayer, F.M, 2024. Generalized βARMA model for double bounded time series forecasting. International Journal of Forecasting, 40, 721–734. DOI: 10.1016/j.ijforecast.2023.05.005
[34] Espinheira, P.L. and Silva, A.O., 2020. Residual and influence analysis to a general class of simplex regression. TEST, 29, 523–552. DOI: 10.1007/s11749-019-00665-3